

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: December 9, 2002, 12:24:14 ; Search time 2054 Seconds  
(without alignments)  
2633.541 Million cell updates/sec

Title: US-09-765-034-2  
Perfect score: 1747  
Sequence: 1 MGIWMANATCKNMLAENA.....KSLTFSRMAHELLSPREK 334

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Fgapop 10.0 , Fgapext 0.5  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cg2.1/USPRO/spool/US09765034/runat.05122002.132016.21489/app.query.fasta.1.519  
-DB=EST -QWTF=fastlap -SUPPL=p2n.ref -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human0.cdi -LIST=45  
-DOCLALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09765034.ecgn.1.1.763.frunat.05122002.132016.21489 -NCPU=6 -ICPU=3  
-NO.XLPYX -NO.MMAP -LARGQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=10 -HREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estmu:\*  
4: em\_estcov:\*  
5: em\_estpl:\*  
6: em\_estro:\*  
7: em\_hic:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estl:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	842.5	48.2	877	12	BG402029
2	737	42.2	633	17	AG083174 Pan trogl
3	692	39.6	1101	17	CNS04WV3
4	655	37.5	956	17	CNS028Y4
5	619	35.4	520	9	AL165305
6	536	30.7	1026	17	CNS051MY
7	500	28.6	422	10	BB847918
8	480	27.5	420	10	BB864882
9	478	27.4	589	12	BF196066
10	477	27.3	639	10	AM612141
11	467	26.7	582	10	BE221739
12	426	24.4	636	10	BB645274
13	402	23.0	416	10	BB846608
14	383	21.9	3001	11	AK017378
15	383	21.9	3001	11	AK005013
16	350.5	20.1	1258	11	AF345567
17	342	19.6	1022	17	CNS04W90
18	330	18.9	946	9	AL551903
19	328	18.8	931	9	AL547762
20	327	18.7	1766	11	AK008952
21	325	18.6	904	9	AL532537
22	322.5	18.5	623	14	B0038875
23	322.5	18.5	2146	17	BC027965
24	319.5	18.3	989	17	CNS03C14
25	316	18.1	2020	11	BC013402
26	306.5	17.5	3005	11	AK019478
27	300.5	17.2	1848	11	AK011967
28	289	17.1	2706	11	AK020001
29	298	17.1	638	9	AL675845
30	296	16.9	641	14	B0396255
31	292	16.7	870	9	AL545172
32	292	16.7	885	9	AL546894
33	286.5	16.4	1057	10	BM918711
34	285.5	16.3	1009	10	BB609892
35	285	16.3	907	9	AL552677
36	284	16.3	1017	14	BM922807
37	283.5	16.2	1063	14	BM920158
38	280.5	16.1	937	17	CNS03GSG
39	280.5	16.1	1372	11	AF345568
40	280	16.0	1074	14	BM917063
41	278	15.9	935	9	AL531903
42	276.5	15.8	942	17	CNS0455H
43	275	15.7	861	13	BI768868
44	274.5	15.7	1076	14	BM923028
45	272.5	15.6	1088	13	BM549799

ALIGNMENTS

RESULT 1  
LOCUS BG402029 877 bp mRNA linear EST 12-MAR-2001  
DEFINITION 602466748F1 NIH\_MGC\_75 Homo sapiens cDNA clone IMAGE:4594810 5',  
mRNA sequence.  
ACCESSION BG402029  
VERSION BG402029.1 GI:13295477  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 877)  
NIH-MGC http://mgc.ncl.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apds-remail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
DNA distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1CM136 row: P column: 11  
High quality sequence stop: 542.  
Location/Qualifiers

FEATURES  
Source

1. 877  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4594810"  
/clone\_1lb="NIH MGC 75"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Kidney; Vector: pDNR-LIB (Clontech); Site\_1:  
Site1 (ggccgcctcgcc); Site\_2: Site1 (ggccatataggcc); 5' and  
3' adaptors were used in cloning as follows: 5' adaptor  
sequence: 5'-CACGGCCATTAAGGC-3' and 3' adaptor sequence:  
5'-ATTCAGAGCGCCGCGCCGACATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.65  
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts  
by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA). Note: this is a NIH-MGC Library."

BASE COUNT 198 a 221 c 199 g 259 t

ORIGIN

Alignment Scores:

Pred. No.: 877  
Score: 842.50 Matches: 176  
Percent Similarity: 90.50% Conservative: 5  
Best local Similarity: 88.00% Mismatches: 16  
Query Match: 48.23% Indels: 5  
DB: 12 Gaps: 2

US-09-765-034-2 (1-334) x BG402029 (1-877)

QY 1 MetleuGlYIleMetAlaTrpAsnAlaThrCysAsnTrpLeuAlaAlaGluAla 20  
DB 101 ATGCTGGGATCATGGATGAAATGCAACTGCAAAACCTGGCTGGCAGAGCTGCC 160  
QY 21 LeuGluLysTyrTyrLeuSerIlePheTyrGlyIleGluPheAlaValGlyValLeuGly 40  
DB 161 CTGGAAGAAGTACTACCTTTTCCATTTTATATGAGATGAGTTCGTTGGAGATCCTTGA 220  
QY 41 AsnThrIleValValTyrGlyTyrIlePheSerLeuLysAsnTrpAsnSerSerAsnIle 60  
DB 221 AATACCAATGTTGTTTACGCTACATCTTCTCTGAAAGAACTGGACAGCAGTAATAT 280  
QY 61 TyrLeuPheAsnLeuSerValSerAspLeuAlaPheLeuCysThrLeuProMetLeuIle 80  
DB 281 TATCTCTTAACCTCTGCTGCTGACTTACCTTTTCTGTCACCCCTCCATGCTGATA 340  
QY 81 ArgSerTyrAlaAsnGlyAsnTrpIleTyrGlyAspValLeuCysIleSerAsnArgTyr 100  
DB 341 AGGAGTATATGCAATGCAAAATGATATATGAGACGCTCTGATATACGACACGATAT 400  
QY 101 ValLeuHisAlaAsnLeuTyrThrSerIleLeuPheLeuThrPheIleSerIleAspArg 120  
DB 401 GTGCTTCATGCAACCTATATACAGCATCTTCTTTCACATTTTATACAGCATATGATGA 460  
QY 121 TyrLeuIleIleLysTyrTrpPheArgGluHisLeuLeuGlnLysGluPheAlaIle 140  
DB 461 TACTTGTAATTAAGTATCTTTCCGAGAACCTTTCGCAAAAGAAAGTTGCTATTTT 520  
QY 141 LeuIleSerLeuAlaIleTyrPheValLeuValThrLeuGluLeuLeuProIleLeuProLeu 160  
DB 521 AATCTC---CTGGCCATTGGGT-TTAGTAACTTAGAGTTACTACCATATCTTCCTT 576

QY 161 IleAsnProValIle-ThrAspAsnGlyThrThrCysAsnAspPheAlaSerSerGlyAs 180  
DB 577 ATAAATCCTGTATTAATGACAAATGCGGCC--TGTAAGATTTTGCAGAGCTGGAGA 634  
QY 180 ProAsnTyrAsnLeuIleTyrSerMetCysLeuThrLeuLeuGlyPheLeuIlePro 199  
DB 635 CCCCAACATCCACATC---TTTTCAGATGTGTCTACCTGTGGGCGCTTATCC 689

RESULT 2

AG083174/c 633 bp DNA linear GSS 03-NOV-2001  
LOCUS Pan troglodytes DNA, clone: PTB-080115.R, genomic survey sequence.  
DEFINITION AG083174  
ACCESSION AG083174.1 GI:16634976  
VERSION  
KEYWORDS  
SOURCE

ORGANISM

Pan troglodytes male lymphoblast DNA, clone\_1lb:PTB Chimpanzee Male  
BAC library clone:PTB-080115.R.  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.

REFERENCE

1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
BAC end sequences of library PTB

TITLE

2 (bases 1 to 633)  
Unpublished  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.

REFERENCE

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: chimpanzee@sc.riken.go.jp, URL: http://gpc.riken.go.jp/  
Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENTS

Clones are derived from the chimpanzee BAC library PTB. This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

PRIMERS

Sequencing: M13Rev  
LIBRARY  
Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI

FEATURES

Location/Qualifiers  
1. 633  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-080115.R"

Source

/sex="male"  
/cell\_type="lymphoblast"  
/clone\_1lb="PTB Chimpanzee Male BAC Library"

BASE COUNT

193 a 129 c 149 g 158 t 4 others

ORIGIN

Alignment Scores:  
Pred. No.: 633  
Score: 737.00 Length: 151  
Percent Similarity: 85.79% Matches: 6  
Best local Similarity: 82.51% Conservative: 25  
Query Match: 42.19% Mismatches: 1  
DB: 17 Gaps: 0

US-09-765-034-2 (1-334) x AG083174 (1-633)

QY 145 AlaIleThrVal-LeuValThrLeuGluLeuLeuProIleLeuProLeuIleAsnProVa 164  
DB 599 GCCATTGTGGTTATAGTAACATAGAGTTGCTACCTTCACTTCACTTAAGAAATCTGA 540  
QY 164 LileThrAspAsnGlyThrThrCysAsnAspPheAlaSerSerGlyAspProAsnTyrAs 184  
DB 539 TAAGACTAACAGAGCGCCACCTTAATATATTTTCAAGGTGTGAGACCCCAAAACAA 480

QY 184 nleuiletysermetcysleuthrleuengllypheleuileproleuphevalmetcy 204  
DB 479 CCTATTACAGCATGTGTCTTACACTGTGTGGGTTCTTATCTCTTTTGTATGG 420  
QY 204 sphepethyrtlyrsilealaleupheleuylslnaragsnarglnvalalathral 224  
DB 419 TTTCTTTTAAGCAAGATGTCTCTCTCTGAAACAGAGAAATAGGAGGTTGCGTCG 360  
QY 224 aleuproleuylulysproleuasnleuvalillemetalaValillepserValpr 244  
DB 359 TCTATCCCTTAAAAAGCTTCACACTGTGTGATCATGACAGGTAATCTCTGTGCT 300  
QY 244 ophethyrtlyrsilemetarganvalargilleaseryleuylserleuylserleuyl 264  
DB 299 TTTACACACCTTATCATGTGATGAGGATGTGGGTTCTCTCTCTGGGAGTTGGAA 240  
QY 264 sglnlyrincysrthrglnvalilleasnseryrthlyrilevalthrargproleual 284  
DB 239 GCAGTNTCAAAAGCAGGAAATATCATCTCTTTTACATGTGACAGCGCTTTGGC 180  
QY 284 apheleuanservalilleasnprovalpheleuylpheleuengllyasphilsphargas 304  
DB 179 CTTTATGACAGTGTGATCAACCTGTCTCTATTTCTTTGGAGATCACTCAGGGA 120  
QY 304 pmetleuethsnlnleuvalrghisnphelyseryleuylserpheserleuyltrpal 324  
DB 119 CATGTGTATGATCAACGAGACACACTCAAAATCCTTACATCTTTAGCATGAGC 60  
QY 324 ahlsglu 326  
DB 59 TCTCGAG 53  
RESULT 3  
CNS04VY3 1101 bp DNA linear GSS 24-MAY-2000  
LOCUS Tetradon nigroviridis genome survey sequence T3 end of clone  
DEFINITION 004B20 of library A from Tetradon nigroviridis, genomic survey  
sequence.  
ACCESSION AL309576.1 GI:8216515  
VERSION AL309576  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetradon nigroviridis.  
ORGANISM Tetradon nigroviridis.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetradon.  
1 (bases 1 to 1101)  
Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetradon nigroviridis DNA sequence  
Unpublished  
2 (bases 1 to 1101)  
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetradon nigroviridis  
Unpublished  
3 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000)  
This sequence is a single read and was generated as part of a large  
scale clone end sequencing project of the Tetradon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetradon.  
location/Qualifiers  
1. 1101  
/organism="Tetradon nigroviridis"  
/db\_xref="taxon:99883"

/clone="004B20"  
/note="Genoscope sequence ID : C0AA004DA10A2-end : T3"  
BASE COUNT 225 a 319 c 274 g 274 t 9 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 5,89e-67 Length: 1101  
Score: 692.00 Matches: 143  
Percent Similarity: 65.99% Conservative: 53  
Best Local Similarity: 48.15% Mismatches: 96  
Query Match: 39.61% Indels: 5  
DB: 17 Gaps: 4  
US-09-765-034-2 (1-334) x CNS04VY3 (1-1101)  
QY 19 AlaAlaLeuGlulysrtyrleuSerilepethyrglyleuValValGlyVal 38  
DB 79 GCTTACTGGAAGAGACTCTCTGACACCTGTACGGTGTGAGTTCGCTGGGTTT 138  
QY 39 LeuGlyAsnThrIleValValTyrglyTyrllePheSerleuLysAsnTrpAsnSer 58  
DB 139 CCCGGAACTGTGTGTGTCTGTGTATATATTTGCTGCGAAGTGTGGACAGCTCG 198  
QY 59 AsnIleTyrlleuPheAsnleuSerValSerAspLeuAlaPheleucysthrleuPromet 78  
DB 199 CATATCTACCTCTTCAACCTGGCCGCGTCAGACCTGTCTCTCTGACAGCTGCCGAC 258  
QY 79 LeuIleArgSerTyrlalasnglyAsnTrpIleTyrglyAspValleucysIleSerAsn 98  
DB 259 CTCTCTACCTCTACACCAACACAGTGTGCTAACAGTCCCTCTCTGTGACAGCCAAC 318  
QY 99 ArgTyValleuHlsAlaAsnleuTyrrhrSerIleleuPheleuThrPheIleSerile 118  
DB 319 GCCTACATCTGACAGCTCAACCTCTACCTCTCTGATGTTGGCGACAGC 378  
QY 119 AspArgTyrlleuIleleuTyrrhrPheArgGlnHlsleuengllyslgIuphe 138  
DB 379 GACCGCTACCTCTGTGTGGGACCCGACGCGGAACTGTGTGTGTCGACCGGNGGCG 438  
QY 139 AlaIleleuIleSerleuAlaIleTyrlleuValleuValThrleuGlucleuProleu 158  
DB 439 GCGCTGTGTGTGACGGGCTCAGCTGTGTGTGTAACTGTGAGGTGGCCCCATGATA 498  
QY 159 Proleu---IleasnprovalIlethrAspAsnGlyThrThyCysAsnAspPheIleaser 177  
DB 499 GCGGTGATGTCTCAGAGCACTCGAAGAGCAACTGACACAGTGAAGACTTCTCCAG 558  
QY 178 ---SerGlyAspProAsnTrpAsnleuIleTyrrSerMetCysleuThrleuengllypne 196  
DB 559 CTGAAGGAGAGCTCAAC---AGCTTGTGTACAGCTGTGGACTGACCTGTGAGTGGTAC 615  
QY 197 LeuIleProleuPheValMetCysPhePheTyrrTyrlleAlaLeuPheleuLysGln 216  
DB 616 TTCCTGCTCTCTGCGCTGTGCGGTTTCACCTACCAATCGGACCTGCTCCATGTC 675  
QY 217 ArgAsnArgGlnVal---AlaThrAlaLeuProleuGlulysProleuAsnleuValille 235  
DB 676 CAGGAAGAGGCTCTGACAGGCGCAGCCCTCTTCAAGGGGCGCATCAGGGTGGCT 735  
QY 236 MetAlaValAlillePheSerValPropheThrProTyrrHlsValMetArgasValArg 255  
DB 736 CTAGTCAGACCAATCTGTGTCTCTCAACCCCATACCTGATGTAGGAACATCAAA 795  
QY 256 IleAlaSerArgleuLyseryrplysglnTyrglnCysThrGlnValAlille-Asnse 275  
DB 796 ATCGCTTACGAAGGTTTGGACCGGAGTGGATCATGACGCTCGGATCATCAAGGG 855  
QY 275 rPheTyrlleValThrArgProleuAlaPheleuAsnSerValIleAsnProvalPheTy 295  
DB 856 GCGTACATTAATCAGCCGACCTGTGCTTTGTCACAGGTCATTAACCTGTCTTCA 915  
QY 295 rPheleuengllyAspHlsPheArgAspMetleuMetAsnGlnleuArg 311







Db	239	CTTCGATCTCGACTTGGCTTTCCGTGGACACCCTCCCATCCTATTAAAGAGTATGCC	298
Oy	85	AasnGLysanTPPIleTyGlyAspValIeucYsIIesSeranaTYtyValIeuHisAla	104
Db	239	AATGATAAAGGAGCACTTAAGAGATGTCTCTATTAAACAACCGATATGTGCTTACACC	358
Oy	105	AnnLeuTYrThrsSerTllePheUthrPhrIleSerTlleAspArgTyrIleuIelle	124
Db	359	AACCTCTACACAGCATCCTCTCTCTCACCTTTCATTACATGACCGATATCTGGTCATG	418
Oy	125	Lys	125
Db	419	AAG	421
RESULT 8			
LOCUS	B8664882	420 bp	mRNA linear EST 27-NOV-2001
DEFINITION	B8664882 RIKEN full-length enriched, RCB-1283 B16 melanoma CDNA Mus		
ACCESSION	B8664882		
VERSION	B8664882.1	GI:17111092	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
JOURNAL	1 (bases 1 to 420)		
COMMENT	Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nunaseki,R., Okazaki,Y., Okido,T., Sakai,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watabiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)		
TITLE	Unpublished (2001) Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) Wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanhi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a> ) for further details. e mouse tissues.		
FEATURES			
SOURCE	location/Qualifiers		
	l..420		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="G43004/C11"		

Accession	Source	Length	Score	Percent Similarity	Best Local Similarity	Query Match	DB
US-09-765-034-2 (1-334) x BB864882 (1-420)		420	9.47e-44	480.00	85.00%	74.17%	27.48%
Alignment Scores:		Length:	420	Matches:	89	Conservative:	13
Pred. No.:		Matches:	89	Conservative:	13	Mismatches:	17
Score:		Mismatches:	17	Mismatches:	17	Indels:	1
Percent Similarity:		Indels:	1	Gaps:	0		
Best Local Similarity:		Gaps:	0				
Query Match:							
DB:							
BASE COUNT	102 a	103 c	87 g	128 t			
ORIGIN	<pre> /cell_line="RCB-1283 B16 melanoma" /mclone.lib="RIKEN full-length enriched, RCB-1283 B16 melanoma cDNA" /cell_line="RCB-1283 B16 melanoma" /notes="pooled cell lines : (cell_line=CR1-1751 WEHI 164), (cell_line=CR1-2116 JC), (cell_line=RCB-0035 WEHI-3), (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA), (cell_line=RCB-0559 K-1, F1), (cell_line=RCB-1283 B16 melanoma), (cell_line=B cells, cell_line=CR1-1702 WEHI 231), (cell_line=Leydig cells, cell_line=CR1-2065 MTC-1), (cell_type=Nullipotent stem cell, cell_line=CR1-2070 NE), (tissue_type=Bladder, cell_line=RCB-0544 MBT-2), (tissue_type=Bone marrow, cell_type=stroma cell, cell_line=CR1-2028 SR-4987), (tissue_type=colon, cell_line=RCB-0549 C1e-H3), (tissue_type=Kidney, cell_line=CCU-142 RMG), (tissue_type=Submandibular gland, cell_line=CR1-1734 SCA-9 clone 15), (strain=BALB/C, cell_type=B cells, cell_line=CR1-1669 BC11 Clone 13.20-353 ), (strain=C3H, tissue_type=brain, cell_line=CR1-1443 BC3H1)" </pre>						

## JOURNAL COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNML, send email to: info@image.llnl.gov  
High quality sequence stop: 483.

## FEATURES

## SOURCE

Location/Qualifiers  
1. 589  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3134907"  
/clone\_lib="NCI CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 173 a 125 c 118 g 173 t  
ORIGIN

## Alignment Scores:

Pred. No.: 2.63e-43 Length: 589  
Score: 478.00 Matches: 90  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 27.36% Indels: 0  
Gaps: 0

US-09-765-034-2 (1-334) x BF196066 (1-589)

OY 245 PheTTPProTyrHisValMetArgAsnValArgIleAlaSerArgLeuGlySerTrpLys 264  
|||||  
DB 589 TTTACACCCCTATGACGTCATGCGGATGAGATCGCTTCACGCTGGGAGTTGGAAG 530  
OY 265 GlnTyrGlnCysThrGlnValValIleAsnSerPheTyrIleValThrArgProLeuAla 284  
|||||  
DB 529 CAGTATGATGACACGACGTCGTCATCACTCTTTACATTGTGACACGGCCTTTGGCC 470  
OY 285 PheLeuAsnSerValIleAsnProValPheTyrPheLeuLeuGlyAspHisPheArgAsp 304  
|||||  
DB 469 TTTTGAACAGATGTCATCAACCCCTGCTTCTATTCTTTGGGAGATCAGCTTCAGGAC 410  
OY 305 MetLeuMetAsnGlnLeuArgHisAsnPhelysSerLeuThrSerPheSerArgTrpAla 324  
|||||  
DB 409 ATGCTGATGAATCAACTAGACACAACTTCAATCCCTTACATCCTTAAAGAGATGGCT 350  
OY 325 HisGluLeuLeuSerPheArgGluLys 334  
|||||  
DB 349 CATGAACCTCACTTTCATTCAGAGAAAAG 320

## RESULT 10

AW612141/c

LOCUS AW612141 639 bp mRNA linear EST 23-MAR-2000  
DEFINITION hg94h07.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:2953309 3'

similar to SW:P2YR\_CHICK P34996 P2Y PURINOCEPTOR 1'', mRNA  
sequence.

ACCESSION AW612141 GI:7317327

VERSION AW612141.1

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

## REFERENCE AUTHORS TITLE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 639)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

## JOURNAL COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNML at: [image.llnl.gov/image/html/resources.shtml](http://image.llnl.gov/image/html/resources.shtml)  
Seq primer: -400p from Glibco  
High quality sequence stop: 449.

## FEATURES

## SOURCE

Location/Qualifiers  
1. 639  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2953309"  
/clone\_lib="NCI CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 189 a 131 c 131 g 188 t  
ORIGIN

## Alignment Scores:

Pred. No.: 3.85e-43 Length: 639  
Score: 477.00 Matches: 97  
Percent Similarity: 93.40% Conservative: 2  
Best Local Similarity: 91.51% Mismatches: 6  
Query Match: 27.30% Indels: 1  
Gaps: 0

US-09-765-034-2 (1-334) x AW612141 (1-639)

OY 230 ProLeuAsn-LeuValIleMetAlaValIlePheSerValProPheTrpTyrHis 249  
|||||  
DB 634 CCTGCAATCTGTGGTCATCATGACGACGTGTAATATCTGTCGTATTAACACCATATCA 575  
OY 249 sValMetArgAsnValArgIleAlaSerArgLeuGlySerTrpLysGlnTyrGlnCysThr 269  
|||||  
DB 574 CGTCATGCGGATGATGAGATCGCTTCACGCTGGTGAAGGATGATCGATCGATGAC 515  
OY 269 GlnValValIleAsnSerPheTyrIleValThrArgProLeuAlaPheLeuAsnSerVal 289  
|||||  
DB 514 TCAGTGTGTCATCACTCTTTACATTGTGACACGGCCTTTGGCCTTTGGAACAGTGT 455  
OY 289 IleAsnProValPheTyrPheLeuLeuGlyAspHisPheArgAspMetLeuMetAsnGln 309  
|||||  
DB 454 CATCAACCCCTGCTTCTTCTTATGAGGAGATCAGCTTCAGGACATCGATGAATCA 395  
OY 309 nLeuArgHisAsnPhelysSerLeuThrSerPheSerArgTrpAlaHisGluLeuLeu 329  
|||||  
DB 394 ACTGAGACACAACTTCAATCCCTTAAAGAGATGGCTCATGAATCCTCACT 335  
OY 329 userPheArgGluLys 334  
|||||  
DB 334 TTCAATTCAGAGAAAAG 319



RESULT 11  
BE221739/c 582 bp mRNA linear EST 03-JUL-2000  
LOCUS BE221739  
DEFINITION h158c09.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:3132688 3'  
similar to SW:P2YR\_HUMAN P47900 p2y PURINOCPTOR 1 ;, mRNA  
sequence.  
ACCESSION BE221739.1 GI:8909046  
VERSION BE221739.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 582)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
inf@image.llnl.gov  
Seq primer: -40UP from Gibco  
High quality sequence stop: 479.  
FEATURES  
source  
Location/Qualifiers  
1..582  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3132688"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clones 132376-132911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."  
BASE COUNT 169 a 125 c 118 g 170 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,46e-42 Length: 582  
Score: 467.00 Matches: 88  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 26.73% Indels: 0  
DB: 10 Gaps: 0  
US-09-765-034-2 (1-334) x BE221739 (1-582)  
QY 247 ProTyHisValMetArgAnValArgileAlaSerArgLeuGlySerTrpLysGlnTyr 266  
|||||  
DB 582 CCTATCATCGTCAATGCGGATGTGAGGATCGCTTACACGCGCTTGAGGAGTGTGAGCAGTAT 523  
QY 267 GlnCyThrGlnAlaValIleAsnSerPheTyrlleValThrArgProLeuAlaPheLeu 286  
|||||  
DB 522 CAGTGGACATCAGGTGTCATCAACCTCTTTTACATTTGTGACAGGCTTTGGCTTTCTG 463  
QY 287 AsnSerValIleAsnProValPheTyrrPheLeuGlnGlyAspHisPheArgAspMetLeu 306  
|||||  
DB 462 AACAGGTCTATCAACCTGCTCTTCTATTTCTTTTGGGAGATCAGCTTCAGGAGCATGCTG 403  
QY 307 MetAsnGlnLeuAlaArgHisAsnPhenylSerLeuThrSerPheSerArgTrpAlaHisGln 326

DB 402 ATGAATCAACTGAGACACACTTCAATCCCTTACCTTTCAGGATGGGCTCATGAA 343  
|||||  
QY 327 LeuLeuLeuSerPheArgGlnLys 334  
|||||  
DB 342 CTCCTACTTTCATTCAGAGAAAG 319  
RESULT 12  
BB645274 636 bp mRNA linear EST 31-AUG-2001  
LOCUS BB645274  
DEFINITION BB645274 RIKEN full-length enriched, 4 days neonate male adipose  
Mus musculus cDNA clone B430012021 5', mRNA sequence.  
ACCESSION BB645274  
VERSION BB645274.1 GI:15402306  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 636)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takamashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.c.riken.go.jp,  
URL:<http://genome.gsc.riken.go.jp/>  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa,  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,  
Ishii,Y. and Hayashizaki,Y.  
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.  
Funct. Genomics 2 pre, 172-186 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
e mouse tissues.  
FEATURES  
source  
Location/Qualifiers  
1..636  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="B430012021"  
/clone\_lib="RIKEN full-length enriched, 4 days neonate  
male adipose"  
/sex="male"  
/tissue\_type="adipose"  
/dev\_stage="4 days neonate"



Best Similarity:	80.83%	Mismatches:	12
Pest Local Similarity:	70.83%	Indels:	22
Query Match:	23.01%	Indels:	3
DB:	10	Gaps:	0

US-09-765-034-2 (-1-334) x BB846608 (1-416)

Oy	5	MetaIatRPasAlaThrcyslsYsnrPrleuaIalagLualalaLeugluYstyR	24
Dd	59	ATGGCAGAAATTATTTCTTGATGGAATTGGTGGACAGAGAGCATTCGTGAATAAGTAC	118
Oy	25	TyrLeuSerIlePheTyGlylleGlupheValaIGlyValleUGlyAsnThrIleVal	44
Dd	119	TACCTCTCGCACTTTATGCAATCCAGATTCATTTTTGGACTCTGGGAATCTCACTGG	178
Oy	45	ValTYrgLYTyrIlePheSerLeuLysasnTPasnsSerSasnIleTYrLeuPheaSn	64
Dd	179	GNGTTGGCTACCTCTCTTCGCATGGAAGAAGCTGGAAGACAGCATCTATCTTTTAAA	238
Oy	65	LeuSerValSerAspLeuAlaPheLeucSYstrIleuPrometLeuIleArgSerTYrala	84
Dd	239	CTTTCATC-TCTGCATCTTGCTTCTCTGTCACCCTTCATC-CTGATTAAGAGTATGCC	296
Oy	85	ASngLYsnrTPllETyGLYAspVallEuCYslleSeAsnArGYrValleuHsaA	104
Dd	297	AATGATTAAGGGACCTATGAGATCTTCATGTAATACGAACCGATGCTGTTACAAC	356
Oy	105	AsnLeuTYrThrserIlePheLeuPheuthrPhelIeSer-IleAsPARGYrLeuile	123
Dd	357	AACCTTAACACGACATCTCTCTCTCACCTTCATTACGACATGACACCATCTCTCTC	414

RESULT 14

LOCUS AKO17378 2542 bp mRNA linear HTC 19-JAN-2002

DEFINITION Mus musculus 6 days neonate head cDNA, RIKEN full-length enriched library, clone:5430432J15;putative receptor P2Y, G-protein coupled 2, full insert sequence.

ACCESSION AKO17378

VERSION AKO17378.1 GI:12856588

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) 6 days neonate head cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:s430432J15.

ORGANISM Mus musculus

REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumetazoa; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS 1 Carninci,P. and Hayashizaki,Y

TITLE High-efficiency full-length cDNA cloning

JOURNAL MEDLINE PUBMED 99279253

REFERENCE 2 10349636

AUTHORS 3 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL MEDLINE PUBMED 20499374

REFERENCE 11042159

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsui,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichipillary sequencer

JOURNAL MEDLINE PUBMED 20530913

REFERENCE 11076861

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS
4	Kawai, J., Shinaawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Katsukawa, T., Saito, R., Kado, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, M., Gaasterland, Y., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, T., Pesole, G., Quackenbush, J., Schriml, L. M., Stabill, F., Suzuki, R., Tomita, M., Wagner, E., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Cenciari, P., de Bonaldo, M. F., Brownstein, M. J., But, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, K., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Mochizuki, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schombach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokawa, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlsch, S. and Hayashizaki, Y.	Functional annotation of a full-length mouse cDNA collection	Nature	409 (6821), 685-690 (2001)	21085660	11217851	
		5 (bases 1 to 2542)					
	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., But, C., Cenciari, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hasegawa, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, U., Kojima, Y., Konno, H., Konda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numata, K., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinaawa, A., Shiraki, T., Sobabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamana, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.	Direct Submission					
		Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)					
		Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details.					
		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome science laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGACAGATTCGAGTTAAATTAATGATCCGCCGCCGCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAAATTAATGATCCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pluescript KS(+) after bulk excision from lambdaBlaDc I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.					
		Location/Qualifiers					
		1. 2542					
		/organism="Mus musculus"					
		/strain="C57BL/6J"					
		/db_xref="PANTOM_DB:5430432015"					
		/db_xref="MGD:MG1:1897121"					
		/db_xref="taxon:10090"					
		/clone="5430432015"					
		/tissue_type="head"					

FEATURES	Source

```

gene
CDS
/clone.lib="RIKEN full-length enriched mouse cDNA library"
/dev.strage="6 days neonate"
1..2542
/gene="P2ry2"
484..1605
/note="data source:MGI, source key:MGI:105107,
evidence:ISS
purinergic receptor P2Y, G-protein coupled 2
putative"
/codon.start=1
/protein.id="BAB30719.1"
/db.xref="GI:12856589"
/db.xref="MGI:105107"
/translation="MAADLEPMNSTINGTNGDELGKCRNEDPFKYLPLVSTGVC
VLGLINVALYIFLCRLKTNASTYTNFHLAVSDLSLAALPLVYAYAGDMPFS
TVLCGLVFLFETLVCISLFTICISVHRCISLHSLMWSGARVARARAVVWL
VLACQAVLLEVTTSVSGRTCTCHDTSARELSFHVAVSMGLLFLVPSVILVCY
VLAKARLLKPAKGTGGGLPRAKRSVRTIALVAVLAFAPLPHLTRLTYSPSLDL
SCHTLNAINMAYKTRPLASANSCLDPLVYLAQORLYRFAADAKPRTEPPSPQARR
KLGHRPRKRYKRLSVSSDSRSTESTPACSETKDRL"
BASE COUNT      517 a      737 c      705 g      582 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:      1.12e-31      Length:      2542
Score:          383.00      Matches:      96
Percent Similarity: 50.33%      Conservative: 58
Best Local Similarity: 31.37%      Mismatches: 136
Query Match:    21.92%      Indels:      16
DB:              11      Gaps:        7

US-09-765-034-2 (1-334) x AK017378 (1-2542)
QY 7 TTPASNAATHRCysLys---AsnTRPLEuAlaAlaGluAlaLeuGlu----- 22
DB 505 TGGAAATACACCATGATGACGACCTGGAGGGGAGCAACCTGGATACAGATGCTGTTTC 564
QY 23 -----LysTyr---TyrLeuSerIlePheTyrGlyIleGluPheValGly 37
DB 565 AAGCAGGACTTCAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 624
QY 38 ValLeuGlyAsnThrIleValValTyrGlyTyrIlePheSerLeuLysAsnTyrPasnSer 57
DB 625 TTGTGCTCCGAAAGCTGCTGCTCTCTATATCTCTATGCGCGCTTAAACCTGGAGACGC 684
QY 58 SerAsnIleTyrLeuPheAsnLeuSerValSerAspLeuAlaPheLeuGlyThrLeuPro 77
DB 685 TCCACGACCTACATGTTTCACCTGGACGTTTCGACCTCTCTACGACGCGCTGCGCG 744
QY 78 MetLeuIleArgSerTyrAlaAsnGly---AsnTyrIleTyrGlyAspValLeuGly 96
DB 745 CTGTGGGTTTATTACTAGCCCGGCGGAGTGCACCTGCTTACGACGCGTCTGCAAG 804
QY 97 SerAsnArgTyrValLeuHisAlaAsnLeuTyrThrSerIleLeuPheLeuThrPheIle 116
DB 805 CTGGTGGCTTTCTCTTTATACCAACCTCTATGTCAGATCTCTTCTCACCCTGCATC 864
QY 117 SerIleAspArgTyrLeuIleLeuTyrTyrProPheArgLysHisLeuLeuGlnLys 136
DB 865 AGCGGACGCGGCTGCTGCGGAGTCTGCGCCCTGCACCTCCCTGCGTGGGCGCGCGCC 924
QY 137 GluPheAlaIleLeuIleSerLeuAlaIleTyrPValLeuValThrLeuGlnLeuLeuPro 156
DB 925 CGTATATGCCCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 984
QY 157 IleLeuProLeuIleAsnProValIleThrAspAsnGlyThrThrCysAsnAspPheAla 176
DB 985 GTGCTTACTTCGTCACACACGCGTGGGGAACCCGGATCATTGCGCATGACCTCG 1044
QY 177 SerSerGlyAspProAsnTyrAsnLeuIleTyrSerMetCysLeuThrLeuLeuGlyPhe 196
DB 1045 GCCGAGAGAGCTTTTACGCAATTTTGTGCTTACACGCTCCGTCATGCTGCGTCTGCTTTT 1104

```

```

QY 197 LeuIleProLeuPheValMetCysPhePheTyrTyrIleAlaLeuPheLeuLysGln 216
DB 1105 GCTGTGCTCTTTCCGTTATCTGCTGATCGCTTATGCGCTTATGCGGCGGCTCTCAA 1164
QY 217 ArgAsnArgGlnValAlaThrAlaLeuPro-----LeuGlnLysProLeuAsnLeuVal 234
DB 1165 CCGGCTTATGGACACACAGAGCTGCTGCTGCGGCCAACCGCAATCTGTGGCACCAT 1224
QY 235 IleMetAlaValAlaIlePheSerValProPheThrProTyrHisValMetArgAsnVal 254
DB 1225 GCGTTGATGATGCGCGCTTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1284
QY 255 ArgIleAlaSerArgLeuGlySerTyrPysGlnTyrGlnCys---ThrGlnValValIle 273
DB 1285 TACTACTCTCTCCGATCA-----CTTGACCTCAGCTGCGACACCCCTCAACGCCATC 1335
QY 274 AsnSerPheTyrIleValThrArgProLeuAlaPheLeuAsnSerValIleAsnProVal 293
DB 1336 AACATGGCATATATACATACCGCGCGCTGCGGACCGCCACACGCAACAGATTGCTTGA 1395
QY 294 PheTyrPheLeuLeuGly 299
DB 1396 CTCTACTCTCTGCGCAGGG 1413

```

```

RESULT 15
AK005013
LOCUS
DEFINITION
Mus musculus adult male liver cDNA, RIKEN full-length enriched
library, clone:1300015C04;purinergic receptor P2Y, G-protein
coupled 2, full insert sequence.
ACCESSION
AK005013.1 GI:1283638
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cDNA library
clone:1300015C04.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishi, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsuno, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,

```



